Some notes on spectrum analysis

Milind Diwan

10/27/2017

Context

- In analysis of spectra in typical particle physics experiment, techniques of covariance are often employed.
- The intention of these notes is to provide the mathematical tools that go into such analysis without the heaviness of a course in the subject.
- Most of this is known, but it is useful to have the basic structure for reference.
- We will develop these tools with an eye towards a reactor neutrino experiment such as PROSPECT or beam experiment such as DUNE
- In such an experiment, a source (which has some parameters) produces a spectrum which is detected in multiple detectors arranged at several different distances. Each detector is characterized by a set of parameters such as position, size, resolution, etc. Lastly, there are a set of parameters that are independent of the detector such as the interaction cross section.
- The spectra observed in each detector must be analyzed with full understanding of these dependences.

Characteristic function method

A characteristic function is a Fourier transform of a probability density function (PDF).

It makes combinations of probabilities easier to calculate and understand.

X is a continuous random variable with probability density function $P(x)$ then the characteristic function is

$$\varphi_X(k) = \int_{-\infty}^{\infty} P(x) e^{ikx}\, dx$$

This allows easy way to generate moments of the PDF.

$\varphi(k=0) = 1$ since it is the integral of the PDF.

$$\langle x \rangle = -i \frac{\partial \varphi}{\partial k}(k=0)$$

$$\langle x^2 \rangle = -\frac{\partial^2 \varphi}{\partial k^2}(k=0)$$

If X and Y are two random variables and $z = f(x,y)$ then the Characteristic function for Z is

$$\varphi_Z(k) = \iint e^{ikf(x,y)} P(x)dx\, Q(y)dy$$

To get the moments of $f(x,y)$ often it is not necessary to evaluate the integral.

e.g. $f(x,y) = x + y \Rightarrow \varphi_Z(k) = \varphi_X(k).\varphi_Y(k)$ .... leave it for you to prove this

Obviously we can expand

$$\varphi_X(k) = 1 + \sum_{n=1}^{N} \frac{m_n}{n!}(ik)^n + O(k^N)$$

where m_n are moments of the PDF about 0. If we take the logarithm of the characteristic function then we can expand it. This is called cumulant generator.

$$Log(\varphi_X(k)) = \sum_{n=1}^{N} \frac{\lambda_n}{n!}(ik)^n + O(k^N)$$

The λ_n are called the cumulants of the PDF. Cumulants and moments are related. It is easier to work with cumulants sometimes.

$m_1 = \lambda_1$ the mean

$m_2 = \lambda_2 + \lambda_1 m_1$ λ_2 is the variance

$m_3 = \lambda_3 + 2\lambda_2 m_1 + \lambda_1 m_2$ $\lambda_3 = \langle (X - m_1)^3 \rangle$ or third central moment

Higher order cumulants do not have simple explanations

Imagine X, Y are independent random numbers then obviously

$$Log[\varphi_{X+Y}(t)] = Log[\varphi_X(t)] + Log[\varphi_Y(t)]$$

With a little thought one concludes that: the cumulants of the PDF of (X+Y) are the sum of the cumulants for X and Y. i.e. the mean of X+Y is the sum of the means of X and Y. The variance of X+Y is the sum of the variance of X and Y, and so on for any order of cumulants.

Gaussian PDF and its characteristic func.

By definition

$$P(x_1,x_2,...,x_n)=Ne^{-\sum_{i\le j}^{n}\frac{1}{2}a_{ij}(x_i-\mu_i)(x_j-\mu_j)}$$

Gaussian multivariate PDF with a mean of μ_i for all x_i.

$N=(Det[2\pi a_{ij}])^{-\frac{1}{2}}$ is the normalization.

The matrix a_{ij} has to be positive definite.

Define $b_{ij}=Inverse[a_{ij}]$ $\rightarrow$ this is the covariance matrix.

let $X=\{x_1,x_2,...\}$ and $K=\{k_1,k_2,...\}$ for short-hand.

$$\varphi(k_1,k_2,...,k_n)=\int_{-\infty}^{\infty}dX\cdot P(X)e^{iK\cdot X}=e^{i\sum_{j=1}^{n}k_j\mu_j}e^{-\sum_{i\le j}^{n}\frac{1}{2}b_{ij}k_ik_j}$$ is the characteristic function

The Fourier transform of a Gaussian yields a "Gaussian function" with $b_{ij}=[a_{ij}]^{-1}$

more about Gaussian

Start with the characteristic function of a multi-variate Gaussian

$$\varphi(K) = e^{i\sum_{j=1}^{n} k_j \mu_j} e^{-\sum_{i \le j}^{n} \frac{1}{2} b_{ij} k_i k_j}$$

The moments of the probability density can be obtained by differentiation

$\varphi(K=0)=1$ since it is the integral of the PDF.

$$-i\frac{\partial \varphi}{\partial k_i}(K=0) = \left\langle x_i \right\rangle = \mu_i$$

$$-\frac{\partial^2 \varphi}{\partial k_i \, \partial k_j}(K=0) = \left\langle x_i x_j \right\rangle = \frac{1}{2} b_{ij} + \mu_i \mu_j \text{ for } i \ne j \quad \text{and } \left\langle x_i^2 \right\rangle = b_{ii} + \mu_i^2 \text{ for } i = j$$

General rule for any moment is $\left\langle x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \right\rangle = (-i)^{\alpha_1+\alpha_2+\ldots+\alpha_n} \left. \frac{\partial^{\alpha_1+\alpha_2+\ldots+\alpha_n} \varphi(k)}{\partial k_1^{\alpha_1} \, \partial k_2^{\alpha_2} \ldots \partial k_n^{\alpha_n}} \right|_{k=0}$

The cumulant generator for a Gaussian shows

$$Log[\varphi(k_1, k_2, \ldots, k_n)] = i\sum_{j=1}^{n} k_j \mu_j - \frac{1}{2} \sum_{i \le j}^{n} b_{ij} k_i k_j$$

that the Gaussian PDF does not have third or higher order cumulants.

Some simple rules and bivariate Gaussian.

If X and Y are two random variables and $z = f(x,y)$ then the Characteristic function for Z is

$$\varphi_Z(k) = \iint e^{ikf(x,y)} P(x)dx\, Q(y)dy$$

$\varphi_Z(k)$ is also known as the expectation value $E[e^{ikf(x,y)}]$

$$\varphi_Z(k) = \varphi_X(k).\varphi_Y(k) \rightarrow \ Log[\varphi_Z(k)] = Log[\varphi_X(k)] + Log[\varphi_Y(k)]$$

Now we setup explicit formula for bi-variate Gaussian to understand relations

$$Log[\varphi_{XY}(t_1,t_2)] = it_1\mu_X + it_2\mu_Y - \frac{1}{2}(\sigma_X^2 t_1^2 + 2\rho\sigma_X\sigma_Y t_1 t_2 + \sigma_Y^2 t_2^2)$$

$$P_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} Exp[-\frac{1}{2(1-\rho^2)} \times \left[x-\mu_X, y-\mu_Y\right].\begin{Bmatrix} \frac{1}{\sigma_X^2} & -\frac{\rho}{\sigma_X\sigma_Y} \\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{Bmatrix}.\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}]$$

$$P_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} Exp[-\frac{1}{2(1-\rho^2)} \times \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho}{\sigma_X\sigma_Y}(x-\mu_X)(y-\mu_Y)\right]$$

Notice that if the covariance matrix is not full rank (rows are not indep.) then there is no PDF. There can be confusion over random variables that are not-independent versus those that are degenerate. Not independent does not mean degenerate.

Linear combinations

X and Y are indep Gaussian random numbers and

$Z = aX + bY$, $S = \alpha X + \beta Y$. What is the joint char. function for Z and S ?

$$\varphi_{ZS}(t_1,t_2) = \int e^{it_1(ax+by)} e^{it_2(\alpha x+\beta y)} P_{XY}(x,y)dxdy$$

$$\varphi_{ZS}(t_1,t_2) = \int e^{i(at_1+\alpha t_2)x} e^{i(bt_1+\beta t_2)y} P_{XY}(x,y)dxdy$$

$$\varphi_{ZS}(t_1,t_2) = \varphi_X(at_1+\alpha t_2).\varphi_Y(bt_1+\beta t_2)$$

Take the logarithm and collect the terms to get the covariance of Z, S

$$\mathrm{Log}[\varphi_{ZS}(t_1,t_2)] = i(at_1+\alpha t_2)\mu_X + i(bt_1+\beta t_2)\mu_Y - \frac{\sigma_X^2}{2}(at_1+\alpha t_2)^2 - \frac{\sigma_Y^2}{2}(bt_1+\beta t_2)^2$$

$$= i\mu_Z t_1 + i\mu_S t_2 - \frac{\sigma_Z^2}{2}t_1^2 - \frac{1}{2}(2\rho_{ZS}\sigma_Z\sigma_S)t_1t_2 - \frac{\sigma_S^2}{2}t_2^2$$

From this one can just pick out the powers of t for the cumulants (or moments).

$$\mu_Z = a\mu_X + b\mu_Y \quad \mu_S = \alpha\mu_X + \beta\mu_Y$$

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 \quad \sigma_S^2 = \alpha^2\sigma_X^2 + \beta^2\sigma_Y^2$$

$$\rho_{ZS} = \frac{\sigma_X^2 a\alpha + \sigma_Y^2 b\beta}{\sigma_Z\sigma_S}$$

Generalize for correlated random vars.

Let X and Y be random variables with joint Gaussian PDF.

$$Log[\varphi_{XY}(t_1,t_2)] = i\mu_X t_1 + i\mu_Y t_2 - \frac{1}{2}(\sigma_X^2 t_1^2 + 2\rho_{XY}\sigma_X\sigma_Y t_1 t_2 + \sigma_Y^2 t_2^2)$$

Put this in matrix form

$$= \mathrm{i}\begin{bmatrix} \mu_X & \mu_Y \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} t_1 & t_2 \end{bmatrix}\begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$

Let $Z = aX + bY$ and $S = \alpha X + \beta Y$. The characteristic function is easily made

$$Log[\varphi_{ZS}(t_1,t_2)] = Log[\varphi_{XY}(at_1 + \alpha t_2, bt_1 + \beta t_2)]$$

The result is better expressed in matrix form

$$\begin{bmatrix} Z \\ S \end{bmatrix} = \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix} \Rightarrow \begin{bmatrix} \mu_Z \\ \mu_S \end{bmatrix} = \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix}\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

transpose

$$\begin{bmatrix} \sigma_Z^2 & \rho_{ZS}\sigma_Z\sigma_S \\ \rho_{ZS}\sigma_Z\sigma_S & \sigma_S^2 \end{bmatrix} = \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix}\begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\begin{bmatrix} a & \alpha \\ b & \beta \end{bmatrix}$$

example

Y_1 and Y_2 are jointly Gaussian. N is an additional Gaussian r.v.

$E[Y_1]=1,\ E[Y_2]=-1,\ E[N]=0$

$\{Var[Y_1]=4,\ Var[Y_2]=1,\ \rho_{12}=1/2,\ Var[N]=2\}$

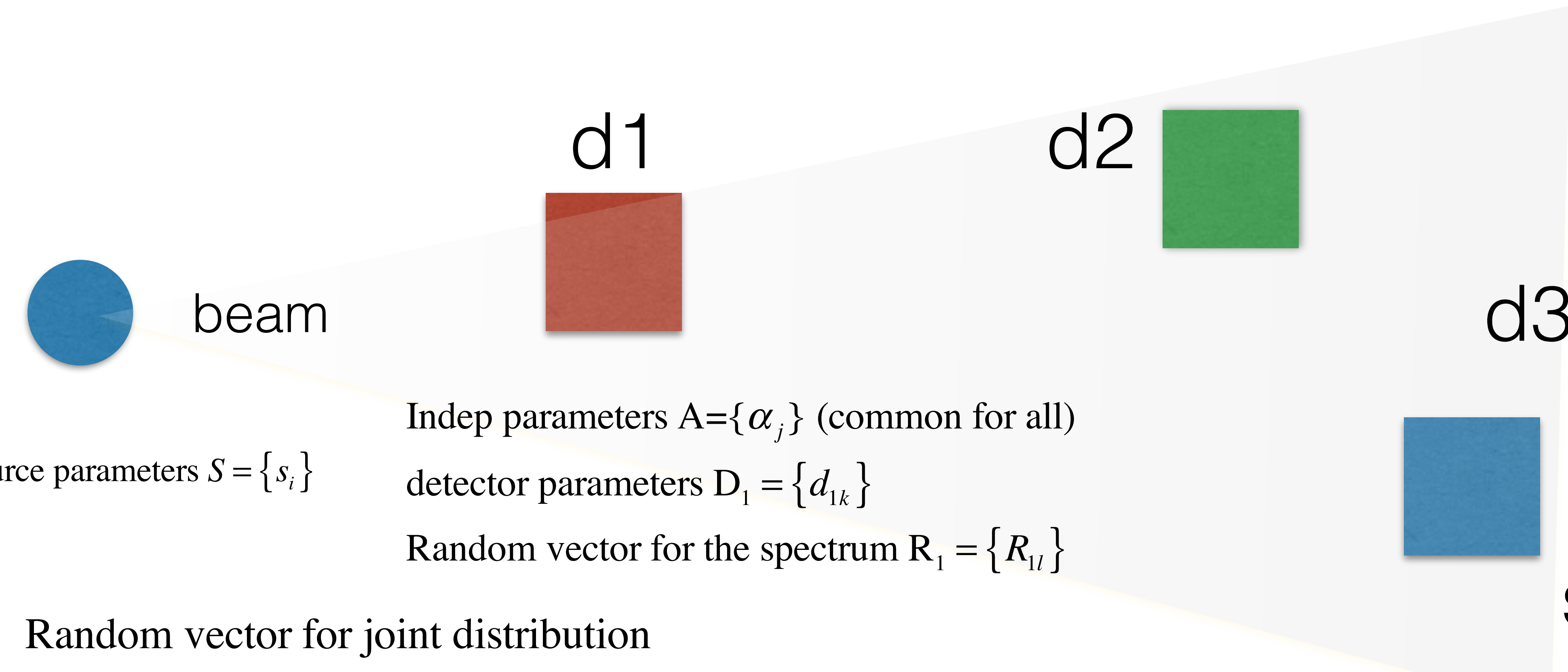
$$Log[\varphi_{12N}(T)]=iT'.\left\{\begin{array}{c}1\\-1\\0\end{array}\right\}-\frac{1}{2}T'\left\{\begin{array}{ccc}4&1&0\\1&1&0\\0&0&2\end{array}\right\}T$$

$X=Y_1-Y_2+N$ and $Z\ =\ Y_1$ what is the joint char. func. of X and Z

$$\left\{\begin{array}{c}E[X]\\E[Z]\end{array}\right\}=\left\{\begin{array}{ccc}1&-1&1\\1&0&0\end{array}\right\}\left\{\begin{array}{c}1\\-1\\0\end{array}\right\}=\left\{\begin{array}{c}2\\1\end{array}\right\}$$

$$\left\{\begin{array}{cc}\sigma_X^2&\rho_{XZ}\sigma_X\sigma_Z\\\rho_{XZ}\sigma_X\sigma_Z&\sigma_Z^2\end{array}\right\}=\left\{\begin{array}{ccc}1&-1&1\\1&0&0\end{array}\right\}\left\{\begin{array}{ccc}4&1&0\\1&1&0\\0&0&2\end{array}\right\}\left\{\begin{array}{cc}1&1\\-1&0\\1&0\end{array}\right\}=\left\{\begin{array}{cc}5&3\\3&4\end{array}\right\}$$

basic analysis structure and nomenclature



Indep parameters A=$\{\alpha_j\}$ (common for all)

detector parameters $D_1 = \{d_{1k}\}$

Random vector for the spectrum $R_1 = \{R_{1l}\}$

Source parameters $S = \{s_i\}$

similar parameters for other detectors

Random vector for joint distribution

$R = \{R_1, R_2, R_3\}$ with mean $\{\mu_1, \mu_2, \mu_2\}$

Block-wise Cavariance is

$$\begin{Bmatrix} \Sigma_1 & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12} & \Sigma_2 & \Sigma_{23} \\ \Sigma_{13} & \Sigma_{23} & \Sigma_3 \end{Bmatrix}$$ where each Σ is a matrix of appropriate dimension

We now have to understand what it means if we measure one of the R_m vectors and ask what the conditional PDF is for the others.

Imagine the neutrino spectrum is measured in D1 and we now want to make a prediction for 2 and 3 given our understanding of the parameters.

Basically, we are seeking the conditional probability for R_2 and R_3 given $R_1 = r_1$

some basics regarding conditional probability

Let $P_{XY}(x,y)$ be a joint probability density function for random variables $X,Y \in \mathbb{R}$. $\varphi_{XY}(t_1,t_2)$ is the joint char. func.
The marginal probability density for X is

$$P_X(x) = \int_{-\infty}^{\infty} P_{XY}(x,y)\,dy$$

The characteristic equation for marginal probability density is obtained by setting the appropriate variables to 0.

$\varphi_X(t_1) = \varphi_{XY}(t_1,0)$ This can be inverted to obtain $P_X(x)$

What is the conditional probability density for Y *given* $X = x$?

$$P_{Y|X=x}(y) = \frac{P_{XY}(X=x,y)}{P_X(x)} \quad \text{if } P_X(x) > 0$$

What is the conditional expectation of Y ?

$$E[Y \mid X = x] = \int_{-\infty}^{\infty} y P_{Y|X=x}(y)\,dy$$

Notice that the conditional expectation for Y now becomes a random variable.

$$E[E[Y \mid X = x]] = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y P_{Y|X=x}(y)\,dy\right) P_X(x)dx = \int y P_{XY}(x,y)\,dx\,dy = E[Y]$$

This is the law of total expectation. similarly we can show

$$Var[Y] = E[Var[Y \mid X = x]] + Var[E[Y \mid X = x]]$$

Total variance of Y is the sum of the "mean" variance under the condition X=x and the variance of the conditional expection of Y.

Let's do this for bi-variate Gaussian

$$Log[\varphi_{XY}(t_1,t_2)] = it_1\mu_X + it_2\mu_Y - \frac{1}{2}(\sigma_X^2 t_1^2 + 2\rho_{XY}\sigma_X\sigma_Y t_1 t_2 + \sigma_Y^2 t_2^2)$$

$$P_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} Exp[-\frac{1}{2(1-\rho^2)} \times [x-\mu_X, y-\mu_Y]. \left\{ \begin{array}{cc} \frac{1}{\sigma_X^2} & -\frac{\rho}{\sigma_X\sigma_Y} \\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{array} \right\} . \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}]$$

For marginal PDF of y

$$Log[\varphi_Y(t_2)] = Log[\varphi_{XY}(0,t_2)] = it_2\mu_Y - \frac{1}{2}(\sigma_Y^2 t_2^2)$$

$$\Rightarrow P_Y(y) = N(\mu_Y, \sigma_Y^2) = \frac{1}{\sqrt{2\pi}\sigma_Y} Exp[-\frac{(y-\mu_Y)}{2\sigma_Y^2}] \quad .... \quad \text{N(,) stands for normal PDF}$$

Conditional probability for X *given* $Y = y$. (I do not know any tricks to do this)

$$P_{X|Y=y}(x) = \frac{P_{XY}(x,y)}{P_Y(y)} = \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} Exp[-\frac{(x-(\mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y-\mu_Y)))^2}{2\sigma_X^2(1-\rho^2)}] = N(\mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y-\mu_Y), \sigma_X^2(1-\rho^2))$$

The conditional expectation is $\mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y-\mu_Y)$

The conditional variance is $\sigma_X^2(1-\rho^2)$ This does not depend on y.

Special cases: The conditional expectation of x is linear with the variation of y from its mean.

If both variables have the same variance and correlation $\rightarrow$ 1 then conditional expectation of $x \rightarrow \mu_X + (y-\mu_Y)$ with variance $\rightarrow 0$. But notice there is then no density.

The multivariate case

Let random vector $R=\{R_1,R_2\}$ with dimensions $n\ and\ m$

The means are given by the vector $\{\mu_1,\mu_2\}$ block-wise covariance is

$$\Sigma=\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad \text{this has sizes} \quad \begin{Bmatrix} n\times n & n\times m \\ m\times n & m\times m \end{Bmatrix}$$

$$Log[\varphi_R(\{T_1,T_2\})]=i\{T_1,T_2\}.\begin{Bmatrix} \mu_1 \\ \mu_2 \end{Bmatrix}-\frac{1}{2}\{T_1,T_2\}.\begin{Bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{Bmatrix}.\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix}$$

By analogy Conditional expectation for $R_1\ given\ R_2=A$. These are all vectors.

The conditional expectation $E[R_1 \mid R_2=A]=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(A-\mu_2)$

The conditional co-variance $E[R_1.R_1 \mid R_2=A]-E[R_1 \mid R_2=A]^2=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ This does not depend on A.

Check the dimensions. Some observations:

Knowing that $\mathrm{R}_2=A$ shifts the mean for R_1 by $\Sigma_{12}\Sigma_{22}^{-1}(A-\mu_2)$

$\Sigma_{12}\Sigma_{22}^{-1}$ is known to be the matrix of regression coefficients.

The covariance for R_1 shrinks depending on the correlation matrix.

Some care is needed here: Σ_{11} and Σ_{22} may not be invertible if their rank is <n,m

We will deal with the issue of rank later. For the moment assume that they are.

setup of the problem

Let's examine what the PDF of spectrum vector R means.
First the PDF is not related to the statistics of the events.

The spectrum R results from the process (generation of neutrinos from the source, and subsequent interactions of those events in the detectors) governed by a set of parameters $\{s,\alpha,d\}$. These parameters have a covariance. We will imagine that the parameters are independent of each others.

Each time we perform the experiment, the parameters take on some values according to the covariance of the parameters.

The covariance of the spectra results from the uncertainty of these parameters.
In the limit that we do an infinite number of experiments, we will get variations of spectra according to the probability density function.

In the case of a single experiment, given the outcome of the spectrum in one detector we want the best estimate for other detector spectra. For this we need to know the covariance of the joint PDF that results from the common set of parameters.

example (make it all 3x3 to avoid the problem of rank)

There are 3 indep. random parameters p={a,b,c} with means {0,0,0} and variances 1, and 4,2.

$$\Sigma_{ab} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The spectrum random variables R_1 have 3 elements and depends on {a,b,c}.

R_2 has 3 elements that depend on {a,b,c}. Let's not worry about constant offsets.

$$\mathrm{R}_1 = C_1.p = \begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 0.1 & 0 \\ 0.5 & 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \qquad \mathrm{R}_2 = C_2.p = \begin{bmatrix} 1 & 0.8 & 0 \\ 0.3 & 1 & 0 \\ 0 & 0.15 & 0.1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Since both a *and* b have a mean of 0, $R1$ *and* $R2$ also have a mean of 0

What is the expected mean and variance of R_1 if a realization of $R_2 = r_2 = \begin{bmatrix} 0.1 \\ -2.4 \\ 0.05 \end{bmatrix}$

First we have to build the block-wise co-variance matrix for $\{R_1, R_2\}$

$$C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \text{ and } C^T = \begin{bmatrix} C_1^T & C_2^T \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} C_1\Sigma_{ab}C_1^T & C_1\Sigma_{ab}C_2^T \\ C_2\Sigma_{ab}C_1^T & C_2\Sigma_{ab}C_2^T \end{bmatrix}$$

Remember that the maximum rank of each diagonal block is same as the rank of Σ_{ab}

what happens when there is not enough information in R2 ?

Let's change the previous example so that R_2 does not "measure" one of the parameters

$$R_2 = C_2.p = \begin{bmatrix} 1 & 0.8 & 0 \\ 0.3 & 1 & 0 \\ 0 & 0.15 & 0.1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \rightarrow\rightarrow \begin{bmatrix} 1 & 0.8 & 0 \\ 0.3 & 1 & 0 \\ 0 & 0.15 & \color{red}{0.0} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Now the rows of C_2 are no longer independent and

$$\Sigma_{22} = \begin{bmatrix} 3.56 & 3.5 & 0.48 \\ 3.5 & 4.09 & 0.6 \\ 0.48 & 0.6 & 0.09 \end{bmatrix} \text{ is only semi-positive definite => Det}[\Sigma_{22}] = 0$$

A true inverse of Σ_{22} does not exist. But we actually do not need it. We need the pseudo-inverse defined for a matrix A in this way

$A.A^g.A = A$. Let's not worry about how this calculation is to be done. The answer is

$$E[R_1 \mid R_2 = r_2] = \{2.52, \quad 0.70, \quad -0.59\}$$

$$Var[R_1 \mid R_2 = r_2] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$$

Since there is no information on parameter "c", the third element of R_1 has remaining variance

what happens when there are more parameters than measurements in R2 to constrain them

imagine 4 indep. random parameters p={a,b,c,d} with means {0,0,0,0} and variances 1, and 4,2.

$$\Sigma_{ab} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}$$

This is the normal situation in an experiment where there are parameters that are hard to constrain.

The spectrum random variables R_1 have 3 elements and depends on {a,b,c,d}.

R_2 has 3 elements that depend on {a,b,c,d}. Let's not worry about constant offsets.

$$\mathrm{R}_1 = C_1.p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.4 & 0.1 & 0 & 0.3 \\ 0.5 & 0.5 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \qquad \mathrm{R}_2 = C_2.p = \begin{bmatrix} 1 & 0.8 & 0 & 0 \\ 0.3 & 1 & 0.1 & 0 \\ 0 & 0.15 & 0.3 & 0.9 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

Since both a *and* b have a mean of 0, $R1$ *and* $R2$ also have a mean of 0

What is the expected mean and variance of R_1 if a realization of $R_2 = r_2 = \begin{bmatrix} 0.1 \\ -2.4 \\ 0.05 \end{bmatrix}$

First we have to build the block-wise co-variance matrix for $\{R_1, R_2\}$

$$C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \text{ and } C^T = \begin{bmatrix} C_1^T & C_2^T \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} C_1\Sigma_{ab}C_1^T & C_1\Sigma_{ab}C_2^T \\ C_2\Sigma_{ab}C_1^T & C_2\Sigma_{ab}C_2^T \end{bmatrix}$$

In normal circumstance the rank of both diagonal matrices should be 3.

continued on the example with 4 parameters

Calculate the matrices (they are somewhat different than before)

$$\Sigma_{11} = \begin{bmatrix} 1 & 0.4 & 0.5 \\ 0.4 & 0.74 & 0.4 \\ 0.5 & 0.4 & 1.75 \end{bmatrix} \quad \Sigma_{12} = \begin{bmatrix} 1 & 0.3 & 0 \\ 0.72 & 0.52 & 1.68 \\ 2.1 & 2.25 & 0.6 \end{bmatrix}$$

$$\Sigma_{21} = \begin{bmatrix} 1 & 0.72 & 2.1 \\ 0.3 & 0.52 & 2.25 \\ 0 & 1.68 & 0.6 \end{bmatrix} \quad \Sigma_{22} = \begin{bmatrix} 3.56 & 3.5 & 0.48 \\ 3.5 & 4.11 & 0.66 \\ 0.48 & 0.66 & 5.13 \end{bmatrix}$$

Calculate the matrix of regression coefficients.

$$\Sigma_{\mathrm{Re}} = \Sigma_{12}\Sigma_{22}^{-1} = \begin{bmatrix} 1.29 & -1.02 & 0.011 \\ 0.524 & -0.372 & 0.326 \\ 0.324 & 0.262 & 0.053 \end{bmatrix} \quad \text{Recall } r_2 = \{0.1,\text{-}2.4,0.05\}$$

Conditional expectation for $E[R_1 \mid R_2 = r_2] = \Sigma_{\mathrm{Re}}.r_2 = \{2.58, 0.963, -0.59\}$

Conditional variance $\mathrm{Var}[\mathrm{R}_1 \mid R_2 = r_2] = \begin{bmatrix} 0.02 & -0.012 & 0.10 \\ -0.012 & 0.008 & -0.06 \\ 0.1 & -0.06 & 0.45 \end{bmatrix}$

and so although the errors have shrunk, they are not zero.

Continued work

Basics of bi-linear forms
Estimation of the co-variance matrix
Independence of parameters
Estimation of parameters